Working Papers 1/2013

Integrated measurement of household-level income, wealth and non-monetary well-being in Finland

Veli-Matti Törmälehto, Olli Kannas and Markku Säylä

Tilastokeskus
Statistikcentralen
Statistics Finland

Inquiries:

Veli-Matti Törmälehto, Olli Kannas and Markku Säylä
09 17 341

Principal editor: Faiz Alsuhail

Layout: Hilkka Lehtonen

© 2013 Statistics Finland

The publications in the Working Papers series do not represent Statistics Finland's viewpoints.
The writers are responsible for the contents of the publications.

Quoting is encouraged provided Statistics Finland is acknowledged as the source.

ISSN 2323–1998
= Working Papers
ISBN 978–952–244–449–3

Integrated measurement of household-level income, wealth and non-monetary well-being in Finland[1]

Abstract

Measurement of economic and non-monetary well-being of households has advanced in recent years, following the "beyond GDP" initiatives and increased availability of cross-national data sources. This paper discusses measurement of household wealth and its relation to income poverty and non-monetary deprivation in Finland, using a survey micro data set that integrates household-level information on income, wealth, and living conditions using telephone interviews, registers, and estimation. In addition to national statistics, the data source has been used as the basis for the Finnish contribution to both the EU Statistics on Income and Living Conditions (EU-SILC, Eurostat) and the Household Finance and Consumption Survey (HFCS, European Central Bank). The paper focuses on concepts and methods. We first discuss the definitions of wealth and the related income flows, using the cross-national sources as the reference, and relate these to the operational definition and measurement in the Finnish implementation. We then use the data to adjust the current main EU indicator on at risk of poverty and social exclusion (AROPE) with household wealth, and examine the sensitivity to the wealth and income concepts and wealth-based poverty thresholds. We conclude by discussing the uncertainties related to the various data sources.

Keywords: wealth, income, poverty, multidimensional counting, data integration, statistical matching

1 Revised version of a paper presented at the 32nd General Conference of the International Association for Research in Income and Wealth, Boston, USA, August 5–11, 2012. The authors wish thank Kaisa-Mari Okkonen for helpful comments.

Contents

Appendices

1 Introduction

With the advances in both methodology and available data, there has been a recent surge in interest on multidimensional measurement of households' well-being (Stiglitz et al. 2009; Alkire & Foster 2011). Information on joint distributions of income, wealth, quality of life and other non-monetary dimensions has become increasingly relevant. Moreover, the financial crisis has highlighted the need to measure also asset-based economic welfare, financial fragility, economic insecurity, and vulnerability to poverty.

While the demand for multi-dimensional data has grown, the resources in the National Statistical Institutes to collect such data have not increased in Europe. One solution to meet the demand without increasing data collection costs is a more efficient use of different data sources. In sample surveys, these include mixed-mode designs such as web-based data collection (CAWI), and sub-sampling or follow-up surveys, as well as record linkage from registers. A more controversial data integration method, statistical matching, may also help to fill in some of the data gaps.

This paper presents an example of a multidimensional survey dataset on income, wealth, and living conditions, constructed by combining telephone-interview data (CATI) with record linkage from registers as well as estimation. Our focus is on household wealth, which was mostly imputed and estimated, and we therefore discuss the concept and measurement of household wealth, as well as uncertainties related to the heteregenous data sources. The primary purpose of the paper is to exploit the various dimensions of the data in order to gain new insights of current measures of economic and non-monetary well-being. To that end, we construct a wealth-augmented variant of the headline EU poverty measure, which measures the share of population who are at risk of poverty or social exclusion (AROPE).

The paper is organized as follows. After a concise description of the survey platform and contents, the operational concepts of wealth, as well as the income flows related to that wealth, are discussed in Section 3. We then move on to the empirical measurement and illustration of the techniques related to the estimation of wealth, and in particular statistical matching of deposits. In Section 5, we illustrate the insights wealth would bring to multidimensional poverty measurement by defining a wealth-based poverty measure and adding it as a new dimension into the main EU poverty measure. Before conclusions, Section 6 discusses the uncertainties related to the data sources, and uses multiple imputation to estimate the imputation variance related to statistical matching.

2 Background: the survey platform

The survey platform we use started in 1977 as the Income Distribution Survey (IDS), combining interview data with registers (mainly income) from the very beginning. In 2003, the EU Statistics on Income and Living Conditions (EU-SILC), coordinated by Eurostat, was fully integrated into the IDS. This expanded the living condition sections in the questionnaire considerably, and new domains such as personal health and material deprivation were added to the survey. In 2009, the IDS-SILC platform was further expanded to cover household wealth, resulting from the launch of the wealth-focused Household Finance and Consumption Survey (HFCS) by the European Central Bank (ECB)[1].

The wealth variables were constructed ex-post from registers by estimation, without adding any new questions to the survey. This re-use of existing data brought a new dimension to the data, and the achieved sample size well exceeded what would have been possible in an independent wealth survey. On the downside, although net worth (assets minus liabilities) and most of its components could be constructed to the dataset, entire sections of the HFCS had to be excluded (e.g. inheritances, consumption) and the level of detail reduced substantially[2]. Table 2.1. gives an overview of the contents of the 2009 data collection and the methods used in the various domains. In brief, income and debts mainly were record linked (exactly matched) from tax registers and various other registers, living conditions were interviewed (CATI), and wealth variables were mostly estimated[3]. In addition to the subjective questions, interviews in some domains are needed to comply with the internationally agreed definitions, or to overcome important gaps in the registers. A concise description of the survey and the data integration process can be found in the Appendix.

In the end, the very same sample of 10,989 households included in the 2009 IDS-SILC database serve two national statistics (Income Distribution Statistics, Household Wealth Statistics) and two European statistics (EU-SILC, HFCS)[4]. In addition to national anonymised micro data sets for researchers, the data are transmitted as micro data to both Eurostat and the ECB and further disseminated as part of cross-national research files.

1 Further information on the HFCS is available in: http://www.ecb.int/home/html/researcher_hfcn.en.html
2 For instance, only total outstanding value of mortages was available, while the requirement was to record 3 mortages with many additional details (e.g. whether the mortage had a fixed or variable interest rate).
3 Record linkage of administrative and statistical register data was deterministic, exact matching, based on unique personal identification numbers (PIN).
4 The reference year for the two national statistics and the HFCS is 2009. Eurostat publishes the results using the fieldwork year of the survey, i.e. as EU-SILC 2010.

Table 2.1
Overview of the domains, primary data sources and methods of the 2009 Survey on Income and Living Conditions (IDS 2009 /EU-SILC 2010 / HFCS 2009)

Domain	Primary source	Respondent/Method
Household structure, housing, housing costs, arrears, child care, child care costs, material deprivation, income (supplementary), employment*,detailed labour variables** personal health**	Interviews	Household respondent, * all persons/proxy, ** selected respondent
Income, education, loans, mutual funds	Registers	Exact matching
Quoted shares, vehicles, forest and farm land	Registers	Exact matching, values estimated
Housing wealth, income (imputed rents)	Registers/interviews	HH respondent/exact matching, values estimated/imputed
Unquoted shares, individual voluntary pension plans	Registers, longitudinal registers	Investment income/accumulation method
Deposits	Wealth Survey 2004	Statistical matching

3 The conceptual framework: wealth and related income flows

Next we discuss the scope of household wealth in this survey by describing the target concepts of non-financial assets (real assets), financial assets, and debts. We take a rather practical approach here, and consider the operational definitions of household net worth and related income flows in the Finnish survey and relate these to the operational cross-national concepts of the HFCS and EU-SILC, i.e. the operational definitions used by the European Central Bank (wealth, income) and Eurostat (income)[5]. As our aim is a joint analysis of income and wealth, the income flows related to the wealth components are important.

The operational definitions in the two European surveys are laid out in EU law (regulations) and guidelines for the EU-SILC, and in the definitions of the HFCS variables for both household wealth and income. International conceptual standards for household wealth statistics were not available at the time of data collection, but OECD has recently published guidelines based on an expert group work (OECD 2013a).[6] The second edition of the Canberra Group Report (2011) is the basic reference for defining disposable income and its sub-components in micro statistics. OECD has recently expanded and further developed the existing guidelines into an integrated framework of income, consumption, and wealth (OECD 2013b). In principle, both EU-SILC and HFCS should be compliant with the Canberra income definition. In practise, there are certain differences and somewhat vague operational definitions.

Net worth generates income, and income generates net worth through positive savings. In principle, all accumulated assets should have their counterparts in current income flows. The receipts that arise from the ownership of non-financial and financial assets generally are defined as property income (Canberra Group 2011). In case labour input is involved, flows are recorded in mixed income/self-employment income. Other flows than income include capital transfers[7] (inheritances and gifts) and revaluation of the assets (holding gains and losses).

As a brief summary, the national income concept is in line with that of EU-SILC (and the OECD) while the concept of net worth does not cover all HFCS items due to practical reasons. In this paper, we also consider two income components, which are not included in the cross-national definitions, namely realised capital gains and net imputed rents of owner-occupiers. Both are directly related to households' asset holdings. In net worth, because of measurement problems, the Finnish implementation does not include valuables, private lending, and life insurance contracts other than private pension plans. Moreover, there are certain shortcomings in coverage and identification, even when the HFCS concept is in principle followed.

5 EU-SILC does not measure quantitatively any component of household wealth.
6 See Kavonius & Törmälehto (2010) for a discussion on the SNA and HFCS concepts of wealth.
7 A part of capital transfers may be recorded as inter-household transfers, as gifts or advancements.

3.1 Non-financial assets and related income flows

Non-financial assets, or real assets, in the HFCS comprise the value of main residence (including land beneath the structures), other properties, vehicles, valuables, and self-employment business wealth (Table 3.1.). In the Finnish survey, valuables are not included and business wealth is not comparably measured, but otherwise the concept in principle covers what is required for the HFCS. There are some conceptual deviations, but except for business wealth these are considered insignificant. Identification of certain assets from the registers is potentially more important, and may lead to some mis-classification of wealth items.

In the HFCS, other properties cover real estate and other properties[8] that are further disaggregated to those for own use, for business use, rented or leased to others, and other uses. In the Finnish implementation, the sub-classes cover own use (holiday homes, second homes), and rented or leased to others (investment real estate); moreover, forests and farmland could be classified as other properties for business use[9].

Regarding business assets, the HFCS measures investments in businesses not publicly traded, and splits these to self-employment and non self-employment parts depending on whether a household member works in such as business. The net value of self-employment businesses not publicly traded are included in non-financial assets. In the Finnish implementation, only the estimated net equity of unlisted limited liability companies is available, without the split to self-employment and non-self-employment parts. In the national definition, these are included in financial assets (other equity) while in the ECB dataset they are non-financial assets since

Table 3.1
The operational concept of non-financial wealth and related income flows

ASSETS		INCOME
HFCS (ECB definition)	FI-HFCS (national definition)	FI-SILC/IDS (excl. realised capital gains)
1 Non-financial assets	Income from non-financial assets	Exact matching
1.1 Household's main residence	Household's main residence	Imputed rents (not included in the official definition)
1.2 Other properties	Other dwellings, forests, farm land	Rental income received / self-employment income
1.2.1 for own use	Dwellings(secondary and leisure homes) Dwellings (investment)	No income
1.2.2 rented or leased to others		Actual rents
1.2.3 for business use	Forest and farm land	Self-employment income
1.2.4 other uses		
1.3 Net value of investments in self-employment businesses not publicly traded	(Partly in unquoted shares, under financial assets)	Property income: dividends from non-listed limited liability companies Self-employment income: part-nership, sole proprietorship
1.4 Vehicles	Vehicles	Not measured
1.5 Valuables	Not collected	Not measured

8 The question wording in the HFCS is as follows: " Apart from your house/apartment) (Do you/Does your household) own any (other) properties, such as houses, apartments, garages, offices, hotels, other commercial buildings, farms, land, etc.?". The FI-HFCS covers mainly dwellings, forests, and farm land, but not e.g. garages

9 The breakdown of other properties for Finland is available only in the national version of the data set.

they have been recorded under "value of additional businesses (HD0900)". This difference is not relevant in this paper, but we use the concept of business assets later on, comprising unquoted shares, farm land and forests.

Income from non-financial wealth is recorded under property income, as rental income (actual and imputed) but partly also in dividends and in self-employment income (assets owned by unincorporated enterprises, royalties, interest paid on producer loans). In addition to actual rents from investment dwellings, i.e. those leased to others, the data set also covers net imputed rents, i.e. implicit net rents from owner-occupied dwellings. In the EU-SILC definition, they are measured but the decision to include them in disposable income is still pending. In the HFCS, imputed rents are not included.

Income flow from other dwellings (free-time & secondary) is not measured, apart from realised capital gains in the national concept in case they are sold. The values of vehicles are included in net worth, but income flow from the services of vehicles or other consumer durables are not included as income.

3.2 Financial assets and related income flows

Financial assets include deposits, bonds, mutual funds, shares and private pension plans in both the HFCS and FI-HFCS, while cash is excluded in both sources (Table 3.2.). The HFCS also includes assets in managed accounts, private lending and whole life insurance contracts, which are not covered in the Finnish implementation[10].

Table 3.2
The operational concept of financial assets and related income flows

ASSETS		INCOME
HFCS (ECB definition)	FI-HFCS (national definition)	FI-SILC/IDS (national definition, excl. realised capital gains)
2 Financial assets	Income from financial assets	Income from financial assets
2.1 Deposits	Deposits	Interest received
2.1.1 Sight accounts		
2.1.2 Savings accounts		
2.2 Mutual funds	Mutual funds	Interest received/dividends (from non-growth funds)
2.3 Bonds	Bonds, subscription rights	Interest received
2.4 Shares and other equity		
2.4.1 Shares, publicly traded	Quoted shares	Dividends
2.4.2 Non-self-employment not publicly traded businesses	Unquoted shares (limited liability companies)	Dividends /self-employment income
2.5 Managed accounts	Not measured	Interest/dividends
2.6 Voluntary pension schemes	Voluntary pension schemes	Income from individual private pension plans
2.7 Whole life insurance con-tract	Not measured (saving and investment plans)	Income from individual private plans
2.8 Private lending (amount owed to household)	Not measured	Interest received
2.9 Other assets (e.g. options, futures, index certificates, precious metals, oil and gas leases, future proceeds from a lawsuit or estate that is being settled, royalties, or something else)	Options	Self-employment income (roy-alties)

Value of business assets when owned as an investor or silent partner are financial assets in the HFCS, whereas in the FI-HFCS only the value of unquoted shares of limited liability companies is estimated. In the ECB dataset, these are recorded under non-financial assets but in the national dataset as financial assets (see Section 3.1.). In the FI-HFCS, other financial assets cover derivatives (forwards, futures, options, swaps). Presumably, of the omissions, only whole life insurance would be of some significance. The HFCS breakdowns of asset components are not typically available, e.g. deposits cannot be broken down to sight and savings accounts[11], and mutual funds cannot be disaggregated to money market and equity funds.

Incomes from financial assets cover interest, dividends, and profits from capital investments in an unincorporated business in all definitions. Annuities from private pension plans are recorded as property income in EU-SILC and the Finnish data, while in the HFCS these are recorded together with occupational pension plans. In Finland, occupational pensions are considered as transfers. Realised capital gains from sales of quoted shares (and mutual funds) are a major income component, which in Finland can be record linked from the tax data. These can be considered as property income along with other taxable realised capital gains, which may also include parts related to non-financial assets. Capital gains are not shown in the table because they can be related to almost all asset types (except deposits). While we have included realised capital gains in income, this has a negligible effect on the poverty indicators used in the second part of this paper.

The HFCS income concept is a bit more aggregated but largely includes the same components as EU-SILC, albeit with different names (e.g. ”income from financial investments”) or recording these together with other components (private pensions). In the HFCS, capital gains are mentioned belonging to ”other income source”, so at least conceptually they seem to be included, and without restriction to realised gains[12]. Royalties are included in self-employment income in both sources.

3.3 *Liabilities and related income flows*

In **liabilities**, the HFCS has the basic breakdown into mortgage debt and non-mortgage debt (Table 3.4.). Mortgages are further divided into those related to principal residence and those related to other properties. Non-mortage debt covers debt without real estate collateral, and includes non-collateralised loans (car loans, consumer loans, and instalment loans), private loans from other households, credit card debt, and credit line/overdraft balances. In the FI-HFCS, measurement is based on the

10 Managed accounts are assumed largely to be included in other asset types (shares, bonds etc.). There are certain coverage problems in the Finnish data, e.g. government bonds and foreign shares are not included in the FI-HFCS because they are not available in the registers.

11 Deposits may be in sight accounts or saving accounts, depending on the conditions restricting money withdrawal from the account (time deposits) and whether it can be used to make payments to third parties (sight accounts) or not (saving deposits).

12 The HFCS blueprint question is formulated as follows: ”Did (you/your household) receive any other regular or irregular income from sources other than those I have already recorded, such as capital gains or losses from the sale of assets, prize winnings, insurance settlements, severance payments, lump sum payments upon retirement, premature withdrawal from private insurance schemes or any other sources in (the last 12 months / the last calendar year)?”

Table 3.3
The operational concept of liabilities and related income flows

LIABILITIES		INCOME
HFCS (ECB definition)	FI-HFCS (national definition)	FI-SILC/IDS (national definition)
3 Liabilities		Interest paid (HMR mortgages, business loans)
3.1 HMR Mortgage debt	HMR Mortgage debt	HMR mortgage interest (imputed rents)
3.2 Other property mortgage	Loans taken up for the purpo-se of acquiring income	Interest paid (self-employment income
3.3 Non-collateralised loans	Debts attributable to farming and forestry, as well as to trade and business activities. Student loans Other liabilities	Not included/measured
3.4 Outstanding debts on credit cards	Not measured (included in 3.3 at least partially)	Not included/measured
3.5 Outstanding debt on credit lines and overdraft balances	Not meaured	Not included/measured
3.6 Private loans from other households	Not measured	Not included/measured

purpose of the loan as it is available in the tax registers: housing loans, education loans, and other loans. Housing loans are recorded as being conceptually equivalent to HMR mortgage debt, but all other loans are recorded under non-collateralised loans (possibly including some loans collateralised by properties other than HMR). Credit card debt, credit lines/overdrafts, and private loans are not measured although credit card debt were covered at least partly in other loans[13].

The flows are related to interest repayments, with those on business and investment loans to be deducted from self-employment income and actual rents. Only if imputed rents are included in income, interest repayments on HMR mortgages are deducted. In the operational definitions, interest paid on consumption loans are not deducted but considered as consumption.

13 In Finland, a loan can have many types of collateral including real and financial assets of household, personal collateral by other households (e.g. parents), or by the state (e.g. student loans, mortgages for own home).

4 Measurement of wealth variables

The challenge was to construct the variables on real and financial assets for the sample households, because most of the other variables required for the national and cross-national datasets (income, demographics and labour, housing, debts etc.) could be taken or derived from the IDS-SILC database. For many asset types, ownership was known from the registers, but the values were not. Valuation was then based on external information (quoted shares, vehicles). As an example, detailed information on ownership of quoted shares was available from tax administration. These were then valued using the end-of-year OMX share price data. For some asset types, the flows related to the asset type were the basis of estimation (private pensions, non-quoted shares). For some asset types, not even the flows of income from the asset type were known. Some relatively minor items were ignored (e.g. certificates of participation). The main missing component was deposits, which had to be statistically matched from another survey.

We next discuss two representative examples of the various methods to estimate the wealth components: the value of a household's main residence and deposits. The estimation of private pension plans is explained in the Appendix. Beyond the methodological interest of statistical matching, deposits are potentially important in asset-based poverty measurement in Section 5. We therefore discuss deposits and the uncertainty related to the estimation method in more detail.

4.1 Housing wealth: mean imputation

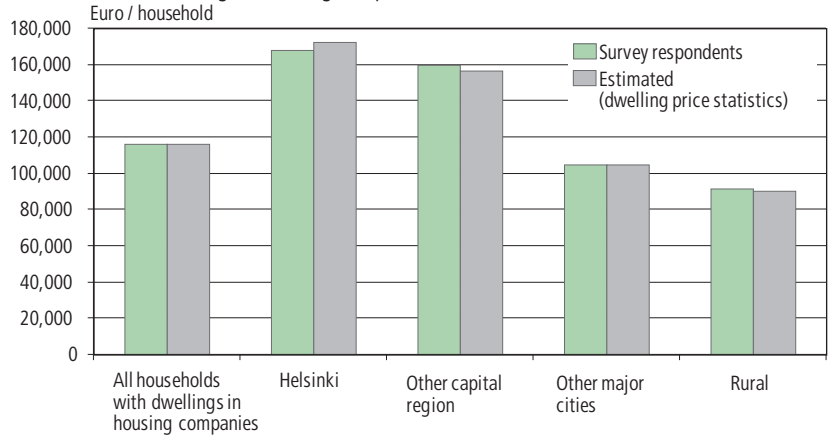
Around two thirds of Finnish households own their dwelling, and the share of total assets was 56 per cent in 2009. Consequently, valid and comparable data on the value of main residence is essential. In wealth surveys, the value of a household's main residence is usually a self-reported estimate of the current market value. This was also the practise in the Finnish surveys prior to 2009, and is the input harmonised method endorsed in the HFCS blueprint questionnaire.

In the 2009 survey, housing wealth was estimated using transaction sale prices. Main residence was identified as the one reported in the survey[14], and then valued based on two methods. For dwellings in housing companies (mostly blocks of flats), purchase prices were linked from asset transfer tax data and deflated to the 2009 value. For other dwelling types, average market prices by strata from dwelling price statistics were multiplied by self-reported floor areas. For both types of valuation, a matching dwelling and its attributes were identified from the register sources for record linkage or to create the strata, controlling for differences e.g. in floor area.

How comparable are the estimated values? Figure 4.1. shows comparison of self-reported values of main residences in housing companies with register-estimated values in the 2004 Wealth survey (see also Kavonius & Törmälehto 2010, 27)[15]. The register-based results and subjective values are quite close to each other. There

14 Secondary and investment dwellings were identified from registers.
15 The comparison is based here entirely on average market prices, i.e. asset transfer tax data are not used.

Figure 4.1
Average value of main residence (euro/household) in 2004, self-reported and estimated (dwelling price statistics) values. Dwellings in housing companies.



Source: Household Wealth Survey 2004

is some supporting evidence from other countries as well: for instance, Cannari and Faiella (2008) examined house prices in Italy based on self-reported values and external sources and concluded that the survey estimates were very close to market values.

4.2 Deposits: statistical matching

No data on households' deposits or any other reliable sources from which the data could be derived (e.g. interest received) are available at the micro level in registers. Interviews offer the only possibility for gathering data on deposits at household level. The most recent data collected by interviews were gathered in the Wealth Survey of 2004. Resorting to such historical data was the only possibility of estimating deposits. The samples of the Wealth survey 2004 and the Income distribution statistics of 2009 cannot be record linked because they contain completely separate sets of individuals. Thus the deposits for the 2009 survey had to be estimated using statistical matching, which is a model-based approach for providing joint information on variables collected through multiple sources for the same populations (Eurostat, 2012).

Statistical matching may be seen as a form of imputation, wherein the values are imputed from the donor sample (Wealth Survey 2004 in this case) to the recipient sample (2009 sample). The 2004 and 2009 samples share a large number of register-based variables (income, loans, education), and the survey part also has many identically defined variables (e.g. household definition, labour, housing). Moreover, the 2004 Wealth survey had essentially the same probability sampling design as in EU-SILC/IDS.

Various matching methods were tested in the project. Euclididan distance function produced adequate results, but was rejected because it contains a subjective element when choosing the weights for the explanatory variables. Regression multiple imputation with random residuals was rejected because it compressed the distributions. The final choice was *unrestricted Hot-Deck nearest-neighbor single imputation*[16] *using predicted values*. Using a pool of shared and comparable covariates found in both the 2004 and 2009 datasets, a regression model was fitted on the 2004 deposits. The predicted values of the regression model without an error term, defined the nearest neighbors in the 2004 and 2009 data sets. Essentially, the imputed value of the i_{th} unit in the recipient file of 2009, is the actual, real observed value in the 2004 donor file, whose predicted value is closest to the predicted value in the 2009 data based on the regression model. The imputed wealth values were scaled uniformly to obtain a closer match with the 2009 level, according to the 43.7 per cent change between 2004 and 2009 in the total sum of household deposits observed in the financial accounts.

Numerous different model specifications were tested before choosing the variables used in the final regression model, which used the log deposits as the dependent variable and asset values, age, income and household characteristics as the explanatory variables[17]. The deposits had mostly weak correlation with other observed variables, and consequently the coefficient of determination of the regression model was only 34.5 per cent. Weak correlations even with income variables (0.23) are most probably due to different asset allocation and risk profiles between the income classes.

The reoccurrence of the same value was not restricted, and the same imputed value from the nearest neighbour in the 2004 survey may occur an unrestricted number of times in the 2009 survey. Restricted matching did not produce as good results as the completely unrestricted imputation model, in particular, because of a low number of observations when matching households within subgroups. The overall sample sizes were very different: 3,211 households in 2004 and 10,989 households in 2009. The use of an unrestricted model makes it possible to always find the actual nearest neighbour for each IDS/SILC household. This is important, in particular because the model's coefficient of determination is relatively low to start with.

Matching method is obliged to presuppose that no significant changes have taken place in the deposit distribution since 2004. This is likely, as structural changes usually take place rather slowly. Without comparison data, however, it is impossible to assess the reliability of the distribution. Nevertheless, the regression model used in this method takes to some extent into account such factors as changes in the demographic and income structure as regards explanatory variables. When such variables as age show a positive correlation with deposits, and the medium age of the population has gone up from 2004 to 2009, the regression model automatically increases

16 A stochastic multiple imputation variant (predictive mean matching) of the chosen method is discussed in the final section of this paper; the final dataset transmitted to the ECB nevertheless includes only one implicate.

17 Housing loans, forest property, vehicles, housing wealth, combined stocks/mutual funds/private pensions and age of the reference person entered as log-transformed continuous variables. Disposable income decile group, socio-economic status (7 classes), household type (5 classes) and tertiary education of the reference person entered as dummy variables.

the average of deposits for the entire population, and thus contributes to changing the deposit distribution as the age distribution changes.

Rässler (2002) provides a review of literature concerning statistical matching, and discusses four levels of validity for statistical matching. The first of these is to check how equal the distributions between the recipient file and the donor file are after the matching procedure. As we can observe from Table 4.1., the matching seems to qualify properly at the aggregate level[18]. Mean, median and the distribution of deposits indicate at least reasonable similarity between recipient file and original sample. In consequence of the unrestricted matching method, standard deviation duly decreases substantially in the imputed sample. The explanatory variables of the matching model naturally show a strong correlation with the deposits in the recipient file. The distributions that are conditional to the background variables are close to the 2004 distribution, with the household income used as an example in Table 4.1.

Table 4.1.
Comparison of basic statistical measures of deposits between donor (2004 Wealth Survey) and recipient (2009 IDS/SILC/HFCS) files (before scaling)

Statistic	Donor file (2004)	Recipient file (2009)
N in sample	3,211	10,989
N in population	2,221,297	2,531,500
Mean € /hh	12,205	13,599
Median €	3,500	4,000
Standard deviation	724,906	476,806
Interquartile range	11,000	12,000
P5	50	60
P10	200	200
P90	30,000	35,000

Income decile	% of total sum, donor	% of total sum, recipient	Mean, donor, €/hh	Mean, recipient, €/hh
I	3.6	2.6	4,403	3,558
II	5.1	4.9	6,196	6,654
III	6.0	5.7	7,269	7,811
IV	9.5	7.4	10,054	10,087
V	9.1	8.7	11,176	11,812
VI	10.0	10.4	11,611	14,102
VII	10.0	11.2	12,148	14,865
VIII	8.2	10.9	12,184	15,247
IX	12.9	14.2	15,679	19,311
X	25.6	23.9	31,299	32,520

The continuous explanatory variables, such as disposable income and wealth variables in the 2004 donor sample, were not scaled to the level of 2009. Obviously, in consequence, regression based imputation explicitly produces monotonically higher imputed deposit values into the recipient file. On this account, it is more meaningful to take notice of the distributions instead of total levels when comparing imputed and original results.

Later, in Section 6, we further validate the statistical matching technique, by estimating the imputation variance of deposits in the total variance of poverty indicators, using stochastic multiple imputation. This is a form of uncertainty analysis, with focus on sensitivity of parameter estimates to the assumptions of the statistical matching model. Another recent approach is the use of Fréchet Bounds (see Leulescu and Agafitaei 2012), which was not attempted here. In terms of Rässler (2002), the sensitivity analysis through multiple imputation falls into checking the second level of validity: preservation of the correlation structure and higher moments of the variables after matching. The third level would be that the true joint distribution of all variables is reflected in the recipient file. The fourth and clearly unattainable level would be that the true unknown values of deposits would be reproduced.

18 We compare the estimated figures of 2009 at their original level before scaling.

5 Wealth and the EU poverty and social exclusion indicator

One of the main benefits of the integrated micro dataset is the possibility to create multidimensional indicators on households' economic well-being, taking into account the joint distributions of the dimensions. In this section, we add wealth as a dimension into the the main EU headline poverty indicator, the population *at risk of poverty or social exclusion* (AROPE). The AROPE-indicator already is multidimensional, combining information on income, employment, and material deprivation. In order to augment this measure with wealth, we follow the multidimensional counting approach proposed by Alkire and Foster (2011). We first focus on the single dimension of wealth, and operationalise an indicator measuring low resources in the wealth dimension. For simplicity and by convention, we call such an indicator an asset-based measure of poverty. We then discuss the left tail of the joint distributions of wealth, income, employment, and material deprivation.

5.1 Wealth and measurement of poverty

Brandolini et al. (2010) and OECD (2013b) provide a synthesis on how wealth can be integrated into measurement of poverty, beyond merely including the income flows from wealth in household disposable income. One alternative would be to combine both income and wealth into a single measure of total resources, a wealth-enlargened income concept[19]. Instead, we follow the approach wherein income and assets are two distinct resource dimensions in a multidimensional poverty measurement framework. This allows us to view assets as a stock of material resources complementary to a flow of current income. Wealth thus offers a *sustainability aspect* to the EU definition of poverty: how long can a household sustain its standard of living by decreasing its accumulated savings, without becoming excluded from the minimum acceptable way of life, should its other resources suddenly fall (e.g. through adverse income shocks)?

To have an operational definition of asset-based poverty, i.e. deprivation in the wealth dimension, we need to define what assets are included and where to draw the asset poverty threshold. A household can be defined as lacking a sufficient buffer of wealth when its wealth holdings are below a certain fraction of income poverty line (Brandolini et al. 2010; Azpitarte 2012). An operational measure depends on the wealth concept, the income concept, the equivalence scale, the income poverty line, and the threshold for wealth. A household is defined as lacking sufficient resources as follows:

$$\frac{W(t)}{e(t)} < \frac{m}{12} * \lambda * median\left(\frac{Y(t)}{e(t)}\right) \tag{5.1}$$

where is a measure of a household's wealth at time point t, e is a measure of consumption units in a household, Y is a measure of the household's annual income,

19 For instance, by replacing the current yield on net worth (property income) with the n-year annuity value of net worth, discounted at rate ρand assuming individuals to evenly spread their wealth in n years (see Brandolini et al., 2010).

λ is a fraction that defines the income poverty threshold, and m is the number of months the household is assumed to have buffer resources for sudden drops in its income. We use the Eurostat scale (modified OECD) as the equivalence scale[20] to define the number of consumption units *e*, and set λ to 0.6, which is the EU standard for relative income poverty (60% of median).

In order to check the sensitivity of the results to the parameters, we vary the asset-poverty threshold (m), the wealth concept (W), and use two alternative definitions of household disposable income (Y): one including imputed rents of owner-occupiers, and one excluding them (cash and near cash income). Imputed rent measures economic benefits of owner-occupied housing (see Törmälehto & Sauli 2013). It is estimated using the rental equivalence approach (market rent of a similar dwelling), net of relevant housing costs and interest repayments on mortgage. Imputed rents reduce the overall level of inequality somewhat, but relative income poverty rate in 2009 remained about the same at 13.1 per cent. There is substantial re-ranking of individuals, however, and imputed rents thus affect the incidence of poverty especially for the elderly (see Appendix 3; Törmälehto & Sauli 2013). Income poverty line in Finland is roughly 10 per cent higher when imputed rents are included, and when asset-poverty is anchored to this, also asset-poverty is necessarily higher.

Regarding the wealth concept, we start with the measure of liquid financial wealth, which here covers quoted shares, mutual funds, bonds, and deposits. This wealth concept may well be the one that would be eventually chosen for descriptive statistical purposes. Since our approach here is a methodological one, we also experiment with alternative wealth definitions. A somewhat less liquid variant is to also add individual private pension plans to liquid financial wealth. These typically can be de-accumulated in case of changes in an individual's situation due to divorce, unemployment or sickness, although there may be significant costs associated to this. These two variants represent "emergency funding" wealth concepts, and it is to be noted that debts have not been deducted.

We further extend the wealth concept by adding non-home real assets net of non-mortgage loans to financial wealth. The loans typically are student loans or loans taken for cars and other consumer durables. Finally, we also show results when all measured net worth is used, with and without business assets (here unquoted shares, forests, and farmland). Table 5.1. shows the asset-based poverty rates according to the various definitions.

The asset-poverty rates appear not to be sensitive to the inclusion of business assets. Because these are not liquid, we exclude them from the wealth concept. Beyond this, it is difficult to justify which concept should be used, although liquid financial wealth is intuitively the appealing option. Our choice is to draw the asset poverty cut-off based on all net worth except net home equity while keeping imputed rents in income. In other words, the value of main residence and the associated loans are not counted as household wealth, but as income flow via imputed rents. Imputed

20 In the Eurostat scale (modified OECD), the first adult gets weight 1, other members aged 14 and over weight 0.5, and those aged 0–13 weight 0.3.

Table 5.1
Asset-based poverty rates (% of population) according to different wealth and income concepts and asset-poverty thresholds (months) in 2009.

	Threshold in months				
	1	3	6	9	12
Disposable income incl. imputed rents					
Asset poverty threshold (equivalent EUR /household)	1,186	3,557	7.114	10,670	14,227
Liquid financial wealth, %	26	46	62	70	75
Liquid financial wealth + private pensions, %	23	42	58	67	72
Net worth excluding net home equity, %	24	32	43	51	58
Net worth excluding business assets, %	17	21	25	28	31
Total net worth, %	16	21	25	28	30
Disposable income excluding imputed rents					
Asset poverty threshold (equivalent EUR /household)	1,077	3,231	6,462	9,692	12,923
Liquid financial wealth, %	24	44	59	68	74
Liquid financial wealth + private pensions, %	21	40	55	64	71
Net worth excluding net home equity, %	23	31	41	49	56
Net worth excluding business assets, %	16	20	25	27	30
Total net worth, %	16	20	24	27	29

Source: Wealth Survey 2009

rents increase the required threshold of wealth holdings and re-rank households in the income distribution by moving mainly homeowners and elderly up in the income ladder. In addition, the chosen concept does not neglect non-mortgage debts, which is the case with liquid financial wealth. As non-mortgage debts are deducted, it is logical to also include their most important asset counterparts (cars and other vehicles, second homes) in household net worth.

Regarding the required buffer, the three-month assumption (25 per cent of income poverty line) seems a reasonable and common compromise. The choice is essentially arbitrary, in a similar way as setting the income poverty threshold to 60 per cent is. One option is to try to estimate the extent of precautionary savings to income. Brandolini et al. (2010) quote several studies on precautionary savings and finally use the three-month assumption (25 per cent) in their own estimates of asset-poverty. In Finland, the income poverty line for a one-person household was EUR 14,227 per year including imputed rents, so the asset poverty line is set to EUR 3,560 per equivalent adult (i.e. 1.5 times this for a two-person household and EUR 7,470 for a household with two adults and two children). Excluding imputed rents, the asset poverty threshold would be EUR 3,230.

5.2 Income poor and asset-based poverty

A common approach in poverty measurement is to focus on the intersection of the two resource dimensions, income and wealth. The table below illustrates how the share of individuals who are both income and asset poor varies when income and wealth concepts change, but fixing the asset-poverty threshold to three months. Following Azpitarte (2012), those who are both income and asset poor are called *twice-poor* while those who are income poor but not asset poor are called *protected (income) poor*. Those who are not income poor but are asset poor are *vulnerable*, i.e. those who currently are not (income) poor but are vulnerable to falling into poverty.

Table 5.2
Income and asset poverty rates (% of population) according to different wealth and income concepts, asset poverty threshold 3 months in 2009.

	(1) Twice-poor	(2) Protected poor	(1+2) Income poor	(3) Vulnerable	(1+3) Asset poor
Disposable income incl. imputed rents					
Liquid financial wealth, %	9.8	3.3	13.1	36.2	46.0
Liquid financial wealth + private pensions, %	9.7	3.4	13.1	32.3	42.0
Net worth excluding net home equity, %	8.2	4.9	13.1	23.8	32.0
Net worth excluding business assets, %	7.3	5.8	13.1	13.7	21.0
Total net worth, %	7.2	5.9	13.1	13.8	21.0
Disposable income excl. imputed rents					
Liquid financial wealth, %	8.9	4.2	13.1	35.1	44.0
Liquid financial wealth + private pensions, %	8.7	4.4	13.1	31.3	40.0
Net worth excluding net home equity, %	7.4	5.7	13.1	23.6	31.0
Net worth excluding business assets, %	5.6	7.5	13.1	14.4	20.0
Total net worth, %	5.6	7.6	13.1	14.4	20.0

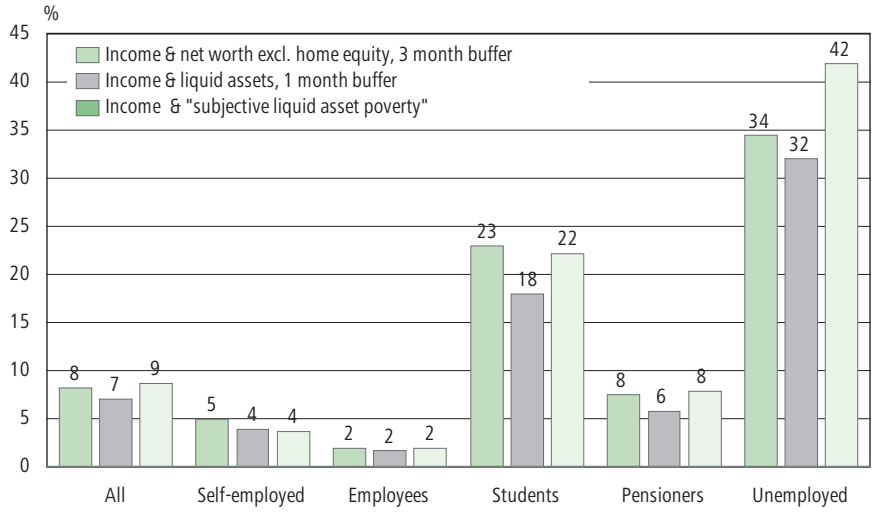
Source: Wealth Survey 2009

The share of twice-poor is 1–1.5 percentage points higher when imputed rents are included in the income concept, and varies around three percentage points depending on the wealth concept. Regarding this, our choice of net worth excluding home equity represents somewhat of a middle ground. The variation in asset poverty rates feeds into variation into those who are considered as vulnerable, varying from around 14 per cent with the most extensive wealth concept to around 35–36 per cent when only narrowly defined liquid assets are considered.

In Appendix 2, we also compare the asset-based poverty measures to a measure obtained with just one survey question, which is asked in all EU-SILC countries. It essentially measures a buffer of liquid (financial) wealth, and has been used as an indicator of "subjective liquid asset poverty" in a cross-country comparison (Morrone et al. 2012). It is based on a household respondent's answer to the following blueprint question: "*Can your household afford an unexpected required expense (amount to be filled) and pay through its own resources?*" In Finland, the amount to be filled was EUR 1,000, independently of the size and structure of the household (i.e. not equivalized). The amount is different in each country, and is determined based on the income poverty threshold. As Appendix 2 shows, the variable would be a reasonable proxy for asset-based poverty, in particular with only liquid financial wealth and a one-month threshold, or net worth excluding net home equity and a three-month threshold.

In Appendix 3, we use the subjective "liquid asset poverty" in the "twice-poor" measure instead of the actual wealth measures. Figure 5.2 shows that overall, both the objective and subjective asset measures lift people out of poverty in a strikingly similar manner, although for the unemployed the subjective measure is signficantly higher. The overall similiarity is an important notion concerning data collection efforts, given the considerable burden to have an objective measure vs. simply using a single survey question, which moreover already exists for all European countries. Unfortunately, it seems that the coherence between subjective and objective asset poverty measures is not as clear-cut in all countries. In a Eurostat study (2013) relying on statistical matching between EU-SILC and the HFCS, the subjective and objective approaches behaved in a similar manner in some countries (Spain, Portugal) but in a different manner in others (Belgium, Italy).

Figure 5.2
Income poor and asset poor by socio-economic status in 2009: objective and subjective (capacity to face unexpected expenses) asset-poverty measures, percentage of population. Disposable income includes net imputed rents.



Source: Wealth Survey 2009/SILC 2010

5.3 The AROPE measure of poverty risk and social exclusion

It is well recognized that poverty is multidimensional (Alkire & Foster 2011). Consequently, poverty in the EU is defined based on both resources and outcomes: *"poor are the persons whose resources (material, cultural and social) are so limited as to exclude them from the minimum acceptable way of life in the Member State to which they belong"* (Council 1985). The main EU indicator on poverty and social exclusion is a multidimensional indicator that measures the share of individuals at risk of poverty or social exclusion (AROPE). The indicator is derived from EU-SILC and aims to measure progress in meeting the inclusive growth goals of the EU 2020 strategy of the European Union.

The AROPE-indicator combines a low-income measure, a composite indicator of severe material deprivation, and a low work intensity indicator. Table 5.3. gives the definitions and estimated shares of individuals in Finland in 2009, along with the

Table 5.3
Share of population at risk of poverty or social exclusion and its sub-components in 2009 (% of individuals).

Indicator	Definition	% of population
At risk of poverty or social exclusion (AROPE)	Income poor or materially deprived or living in a household with very low work intensity	16.3% (+/– 0.9 pp)
Income poverty (AROP)	Equivalent disposable income below 60% of median	13.1% (+/–0.8 pp)
Severe material deprivation	Deprived in at least 4 out of 9 items	2.7% (+/– 0.4 pp)
Very low work intensity	Household members work less than 20% of their capacity	6.9% (+/– 0.6 pp)

Source: SILC 2010

95% confidence interval. The material deprivation indicator is a composite indicator in itself, and consists of nine items. These are arrears (mortgage/rent/utility bills/hire purchase instalments/other loan payments), capacity to afford one week's annual holiday away from home, a meal with meat, chicken, fish every second day, capacity to face unexpected financial expenses, ability to pay for keeping ones home adequately warm, and enforced lack of telephone, colour TV, washing machine or car. Severe material deprivation requires shortfalls in four dimensions out of nine.

According to the AROPE measure, a person is multi-dimensionally poor if she/he is deprived in any of the three dimensions. The AROPE follows the union approach, which requires that sufficiency in every dimension is essential for avoiding poverty. In part because of this, the AROPE appears to be quite inclusive in Finland, and may include people who many would not regard as being excluded from the minimum acceptable way of life in Finland.

While the conventional income poverty indicator is relative to national income levels, severe material deprivation aims to capture absolute differences in European living standards. In Finland and in the Nordic countries in general, it has low prevalence with the current definition. Low work intensity measures exclusion from the labour market, and is defined for households with at least one member aged between 18 to 59 years. It follows that for the elderly only two dimensions define the poverty or exclusion status: low incomes or severe material deprivation.

5.4 *Wealth-augmented AROPE-indicator*

In order to include asset poverty as a dimension in the AROPE, we follow the Alkire-Foster dual cut-off approach to multidimensional poverty measurement (Alkire and Foster 2011). The first step in this method is to define the dimensions and the dimension-specific cut-offs. We use wealth, income and social exclusion as the dimensions and for each apply the cut-off already defined above: 60% of median income, deprived in four out of nine items, working less than 20% of capacity, and possessing net worth (excluding home equity) less than 25% of income poverty line.

The second step is to determine in how many dimensions the person is required to be deprived (k). The AROPE-measure sets the cut-off at k=1, i.e. a union approach. Moreover, to take into account how deprived the deprived are, Alkire and Foster propose to adjust the simple headcounts with the proportion of deprivation suffered by the deprived persons. The dimensions may be weighted, but we use equal weights here.

We define two variants of wealth-augmented AROPE. In the first, we add asset poverty as the fourth dimension into the indicator. In the second, we keep three dimensions but replace income poverty with the twice-poor criterion. Table 5.4. shows both wealth-augmented AROPE measures according to the number of the required dimensions (k). For instance, adding asset poverty and requiring persons to be deprived in at least two of the four dimensions (k=2) would lead to a headcount of 11.2%. The deprived have, on average, 2.416 deprivations out of four. This is 60.4% of the four dimensions, and consquently the dimension-adjusted

Table 5.4
Two wealth-augmented AROPE measures in 2009 by the dimension thresholds, percentage of population.

	1) Asset poverty as the 4th dimension				2) Income and asset poverty *(twice-poor)* as the 3rd dimension		
Dimension cut-off	k=1	k=2	k=3	k=4	k=1	k=2	k=3
AROPE augmented with asset poverty, %	38.7	11.2	3.9	0.7	13.0	4.0	0.7
Average deprivation share, %	35.2	60.4	79.6	100	45.5	72.6	100
Adjusted headcount, %	13.6	6.7	3.1	0.7	5.9	2.9	0.7

Source Wealth Survey 2009 /SILC 2010

headcount would stand at 6.7%. In the extreme case of k=4, the deprived are by definition deprived in all four dimensions, and the unadjusted and adjusted headcounts are the same.

We next compare income poverty and the original AROPE indicator deprivation profiles to the four-dimensional measure with k=2 and the three-dimensional measure with k=1 (Table 5.5.). Comparing to the original AROPE, both variants behave the same way for those who are income poor. The four-dimensional measure also affects people above the income poverty line because it is not a union indicator: only being materially deprived or having low work intensity is no longer a sufficient criterion for being poor.

Table 5.5
Poverty profiles based on wealth-augmented poverty measures in 2009, percentage of persons in the group.

	Income poor	Asset poor	Income and asset poor	Vulnerable	AROPE	W-AROPE 1 (3 dimensions, k=1)	W-AROPE 2 (4 dimensions, k=2)
All	13.1	31.7	8.1	23.6	16.3	13.0	11.2
Socio-economic status							
Self-employed	11.7	25.7	5.0	20.7	12.1	5.8	5.4
Employed	3.1	30.9	2.0	28.9	4.0	3.0	2.6
Students	32.2	43.5	23.0	20.5	35.3	29.3	27.4
Retired	14.0	22.9	7.4	15.5	19.7	14.8	10.9
Unemployed	49.5	52.4	34.4	18.0	62.4	59.5	52.1
Others	15.0	35.7	9.3	26.3	17.7	14.2	12.9
Age							
0–24	18.1	38.5	12.1	26.4	20.4	16.3	15.3
25–34	13.3	43.5	9.3	34.2	15.6	13.5	12.1
35–44	10.2	31.3	5.8	25.5	13.5	11.3	9.7
45–54	9.9	31.1	6.3	24.8	15.1	13.9	11.1
55–64	9.5	21.3	5.4	15.9	15.8	13.2	8.7
65–	12.2	21.2	6.2	15.0	13.1	7.3	6.7
Income level (% of median)							
–40	100.0	69.1	69.1	0.0	100.0	84.8	84.8
0–50	100.0	66.4	66.4	0.0	100.0	82.5	82.5
0–60	100.0	57.3	57.3	0.0	100.0	67.6	67.6
0–70	0.0	44.9	0.0	44.9	14.9	14.9	7.6
70–100	0.0	33.5	0.0	33.5	5.0	5.0	2.1
100–150	0.0	25.0	0.0	25.0	1.3	1.3	0.4
150–200	0.0	13.1	0.0	13.1	0.8	0.8	0.2
200–	0.0	7.4	0.0	7.4	1.0	1.0	0.0

Source Wealth Survey 2009 /SILC 2010

Compared to income poverty, the original AROPE-measure identifies higher risks mainly for those who are not working, but the relative risks do not change dramatically (see Appendix 4). Wealth-augmented AROPE-measures would reduce poverty of the self-employed, while the effect is relatively less pronounced for the unemployed. For these groups, the relative risks also change. Otherwise, wealth-augmentation appears to have little effect on the *relative* deprivation risks although the headcount levels change markedly.

One of the problems of the AROPE measure is that high-income persons can be poor if they do not work or have material deprivation based on subjective responses. As an example, a working-age household may live on its investment income and not be working, and based on the union indicator would be "at risk of poverty or social exclusion". Looking at the poverty profiles by income level, the four-dimensional wealth-augmented AROPE seems to provide some remedy to this.

6 Uncertainty related to the different data sources

The variables used in the wealth augmented AROPE measure come from different sources. Income poverty is mainly determined based on exactly matched register data, while material deprivation and low work intensity are based on interviewed data, and finally asset poverty is determined based on estimated or imputed data.

The joint use of survey and register data affects the total survey error (Groves et al., 2004) essentially by expanding the traditional survey error sources to those related to registers and data integration from multiple sources (see Zhang 2012; Törmälehto & Jäntti 2013). In particular, the *errors related to measurement* (variables) in these sources are different[21]. The interview-based data suffer from traditional survey measurement errors related to validity, measurement, and processing. Regarding measurement, item non-response was limited in the variables used in the AROPE, but other inaccuracies in measurement, validity and processing are hard to quantify. There is some control over validity errors (difference between target concepts and operational measurement), while for the register-based variables (income, loans, mutual funds) the validity errors are the main source of uncertainty, i.e. whether the administrative concept matches the statistical concept or not. This largely is the case here, and it is reasonable to assume that registers contain little measurement errors, and errors related to linking and aligning multiple register data sources are virtually non-existent.

The validation of variables estimated through identification and external valuation was mainly univariate, and based on plausibility of time-series estimates and comparisons to external sources (except for owner-occupied housing discussed earlier). The identification of the ownership from registers was problematic at times and resulted in coverage errors (cars, boats), identification errors in aligned sets, and measurement errors due misclassification (e.g. free-time/secondary dwellings vs. investment dwellings). Valuation was particularly challenging when flows were used to estimate the asset values (unquoted shares, private pensions). The uncertainty related to these methods essentially would require experimentation with other methods or data sources (say, with different external information on prices or with alternative register sources).

Beyond the comparison of empirical distributions of donor and recipient samples, evaluation of quality of statistical matching is challenging (Leulescu and Agafitei 2012). One possibility is to use stochastic multiple imputation in order to estimate the variance due to imputation, see Rubin (1986; 1987). To accomplish this, we multiply imputed the deposits using *predictive mean matching*. In this method, the imputed value for the i_{th} unit in the recipient file is the real observed value of its nearest neighbour in the donor sample drawn randomly from a pool of observed values (nearest neighbours), whose predicted values are closest to the predicted value for the missing value from a simulated regression model, where the parameters are

21 See also Verma & Betti (2010) for a framework and discussion on data accuracy in EU-SILC.

drawn from the Bayesian posterior predictive distribution (Rubin 1986; 1987). The regression model itself is the same as described earlier when discussing deposits.

Consequently, by adding a random component to the model, we create five different statistically matched datasets, all containing one implicate of imputed deposits. We then calculate the indicators taking into account the imputation variance between the five implicates. The between variance (B) is the variance of the point estimates across the *m* implicates:

$$B = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{Q}_i - \hat{Q})^2 \tag{6.1}$$

where $\hat{Q}_i$ is the point estimate from the i_{th} imputed data set, i=1, 2, ..., 5 and the point estimate for Q from multiple imputations is the average of the m complete-data estimates. The within-implicates component of total variance (W) is the average of variances across the m implicates.

Table 6.1. shows selected indicators based on single and multiple imputation. Relatively low between variance for the wealth augmented AROPE indicators and mean deposits / portfolio share of deposits suggests that adding a random component into the matching model has minor impact on the results. This can also be concluded from very low variation of indicators across the implicates. Thus, results from multiple imputation can be interpreted as somewhat of a validation of the efficiency of the matching process and it gives support for using statistically matched information as a part of statistical data and indicators. The matching model seems to qualify surprisingly well, taking into account the low explanatory power of the regression model and the low quantity of observations in the donor file (ratio of observations in the donor file and recipient file was as low as 29 per cent).

Table 6.1
Selected indicators (percentage of population) and their standard errors (percentage points) when deposits are multiply imputed in 2009.

	W-AROPE Income and asset poverty (twice-poor) as the 3rd dimension		W-AROPE Asset poverty as the 4th dimension		Asset poverty		Average portfolio share, % of financial assets		Deposits: mean	
	Per cent	StdErr	Per cent	StdErr	Per cent	StdErr	Mean	StdErr	Mean	StdErr
Single imputation	13.0	0.41	11.2	0.40	31.7	0.59	76.3	0.38	15 396	336
Multiple imputation	13.0	0.45	11.3	0.43	31.5	0.75	75.9	0.40	15 405	406
MI implicate 1	13.1	0.44	11.5	0.40	31.8	0.59	75.8	0.38	15 456	401
MI implicate 2	12.9	0.43	11.1	0.40	31.3	0.59	76.0	0.38	15 388	413
MI implicate 3	12.9	0.44	11.2	0.40	31.4	0.59	75.9	0.38	15 396	367
MI implicate 4	12.9	0.43	11.2	0.39	30.9	0.58	76.0	0.38	15 362	427
MI implicate 5	13.0	0.44	11.3	0.40	32.0	0.60	75.8	0.38	15 424	409
Between-implicates variance	0.0111		0.0215		0.18		0.0145		1 292	
Relative increase in variance	0.08		0.16		0.61		0.1193		0.01	

Source Wealth Survey 2009 /SILC 2010

7 Conclusions and discussion

This paper described the integration of wealth into the 2009 sample of the Finnish Survey on Income and Living Conditions, the conceptual background and measurement issues, and finally used wealth as an additional dimension in a multidimensional poverty measure. The whole survey already was based on data integration of different data sources, and the wealth variables further extended the range of sources and methods. Overall, we conclude that the various estimation methods performed adequately, at least in the univariate sense, and reasonably valid measures of household net worth could be constructed. In some asset types, second best methods had to be used, and we particularly discussed statistical matching of deposits.

As an empirical illustration, we augmented the current EU poverty indicator with an asset-based poverty measure. The wealth-augmented at risk of poverty and social exclusion measure (AROPE) changed levels of poverty, but yielded little surprises in the relative deprivation profiles by lifting the retired and the self-employed out of poverty relatively more. The experiment with a subjective measure of liquid assets proved to yield very similar results with objective wealth measures, at least when the dual condition of being both income and asset poor is considered.

The wealth augmented AROPE is a good example of a highly important multidimensional indicator, wherein all dimensions in our case were based on different data sources: interviews, registers, exact matching, and statistical matching. We verified the sensitivity of the results with respect to statistical matching using stochastic multiple imputation, and conclude that the uncertainty of the poverty indicators related to statistical matching of this specific asset type was quite limited.

A comprehensive measure of wealth cannot be based only on administrative data in Finland. Moreover, a sample survey is necessary also to comply with the demands of the European statistics on the distributions of wealth and income as well as living conditions. The approach presented in this paper is a cost-efficient way of collecting the required multidimensional micro data, and has many advantages in terms of sample size, accuracy, and coherence compared to an independent wealth survey. Nevertheless, there are challenges in the use of multiple data sources for various descriptive and analytical purposes and various national and cross-national users. Conceptual validity and internal consistency of the data have shortcomings; all user needs cannot be met, there are gaps in the measurement, the degree of cross-national comparability is lower, and the approach is vulnerable to changes in the administrative data. The next wave of the Finnish contribution to the Household Finance and Consumption Survey (reference year 2013) will be produced in essentially the same way as the 2009 data used in this paper. This provides a new opportunity to develop and evaluate the methods, and to make a more strategic decision on whether to continue with the integrated approach or to return to the ”stovepipe” model of a separate data collection on household-level wealth.

References

Alkire, S & Foster, J.E (2011). Counting and multidimensional poverty measurement. Journal of Public Economics 95(7–8): 476–487.

Azpitarte, F. (2012). Using both income and wealth: A cross-country comparison between the U.S and Spain. Review of Income and Wealth, number 1, series 58.

Brandolini A. & Magri S. & Smeeding T. M. (2010). Asset-based measurement of poverty. Journal of Policy Analysis and Management, 29.

Canberra Group Handbook on Household Income Statistics, Second Edition 2011. United Nations. New York and Geneva.

Cannari L. & Faiella I. (2008). House prices and housing wealth in Italy. Bank of Italy Working Papers No. A4.

Council (1985). Council Decision of 19 December 1984 on Specific Community Action to Combat Poverty, 85/8/EEC, OJEC, L 2, Brussels.

Eurostat (2012). Data Matching:EU-SILC and HFCS (ECB). Document LC-LEGAL/38/12/EN presented at the 4th meeting of the Task Force on the Revision of EU-SILC Legal Basis, 25–26 October 2012.

Eurostat (2013). Statistical Matching: Joint Statistics on Income, Consumption and Wealth. Document LC-LEGAL/13–2/13/EN presented at the 6th meeting of the Task Force on the Revision of EU-SILC Legal Basis, 16–18 April 2013.

Groves, R. M. & Fowler Jr. & F. J., Couper M. & Lepkowski J.M. & Singer E. & Tourran-Geau, R. (2004). Survey Methodology, New York: Wiley Interscience.

Kavonius, I. & Törmälehto, V-M (2010). Integrating Micro and Macro Accounts – The Linkages between Euro Area Household Wealth Survey and Aggregate Balance Sheets for Households. Paper prepared for the 31st General Conference of the International Association for Research in Income and Wealth, St. Gallen, Switzerland, August 22–28 2010.

Leulescu, A. & Agafitei, M. (2012). A quality framework for matching EU social surveys. Paper presented at the European Conference on Quality in Official Statistics – Q2012, Athens, Greece, 29 May – 1 June 2012.

Morrone A. & Scrivens K. & Smith C. & Balestra, C. (2011). Measuring Vulnerability and Resilience in OECD Countries. Paper presented at the 2011 IARIW-OECD Conference on Economic Insecurity: Measurement, Causes, and Policy Implications, November 22–23 2011, Paris, France.

OECD (2013a). OECD Guidelines for Micro Statistics on Household Wealth, OECD Publishing, Paris.

OECD (2013b). OECD Framework forStatistics on the Distribution of Household Income, Consumption and Wealth, OECD Publishing, Paris.

Rubin, D. B. (1986). Statistical Matching Using File Concatenation With Adjusted Weights and Multiple Imputations. *Journal of Business & Economic Statistics*, 4, 87–95.

Rubin, D.B. (1987). Multiple Imputation for Non-response in Surveys. J. Wiley & Sons, New York.

Rässler. S. (2002). Statistical matching: a frequentist theory, practical applications, and alternative Bayesian approaches. Lecture Notes in Statistics, 168, Springer, New York.

Törmälehto, V-M & Sauli H. (2013). The distributional impact of imputed rent in EU-SILC 2007–2010. Eurostat Methodologies and Working Papers 2013 edition. Luxembourg.

Törmälehto, V-M & Jäntti M. (2013). Combining sample surveys and registers – an overview in the context of EU-SILC. In: Jäntti, M. & Törmälehto, V.-M. (eds.). The use of registers in the context of EU-SILC: challenges and opportunities. Eurostat Methodologies and working papers, The Publications Office of the European Union. Luxembourg.

Stiglitz, J. & Sen, A. & Fitoussi, J-P (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. CMEPSP, September 14 2009.

Verma, V. & Betti, G. (2010). Data accuracy in EU-SILC. In Atkinson, A.B & Marlier, Eric (eds): Income and Living Conditions in Europe. Eurostat Statistical books.

Zhang, Li-Chun (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica* Vol. 66, nr. 1, pp. 41–63.

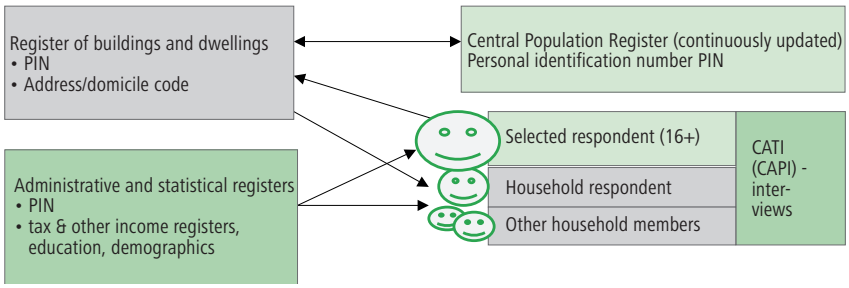
Appendix 1a

Data integration process in the 2009 IDS-SILC database

Figure A1 shows the underlying principle of the integration. Data collection begins from drawing a stratified sample in two phases. A master sample of persons is first drawn from the population database, and all members living at the same address (*co-residents*) with the person selected (*sample person*, *selected respondent*) are then identified and record-linked from the register of buildings and dwellings[22]. To stratify the sample, tax data is then record-linked to all persons before drawing the final sample. High-income households were over-sampled, which is more in line with the ECB requirements (over-sampling of the wealthy).The members of the economic household in the end of 2009 were first inquired in the interview, by correcting the register information on household-dwelling units fed into the electronic questionnaire. For both households and persons, data on the domains described in Table 1 were then collected by interviewing, matching, and estimation/imputation.

The survey was in the field from January to May 2010, and the interviews were mainly CATI-interviews. The response rate was around 70% for the first time interviewed, and the average interview length was around 28 minutes. The questionnaire had little quantitative data, as nearly all income data were linked from registers, so CATI was used as a feasible and cost-effective main mode of collection.Telephone interviewing and the length of the HFCS questionnaire effectively ruled out collecting wealth information for the ECB survey. Wealth data are sensitive and cognitively difficult, so they require face to face interviewing and the associated tools (e.g. cards), as well as thorough interviewer training.

Figure A1 An overview of the data sources.



Depending on the type of question and the analysis unit, the respondent in the interview could be a household respondent (best aware of the financial situation), the selected respondent drawn from the frame, or as an exception a proxy. The subjective questions mainly relate to the time of the interview, but otherwise variables relate mostly to the end of the year (e.g. household composition) or the calendar year. This facilitates register-based checks and editing.

22 Domicile code identifies a household-dwelling unit (household defined only on the basis of co-residence). Personal identification numbers are known for all persons selected from the frame, as well as those residing at the same address.

Record linkage of registers was deterministic, exact matching, based on unique personal identification numbers (PIN). In case of household-level information (e.g. dwellings), linkage was based on the person drawn from the frame. Moreover, all register information were used in micro editing, imputations, and data validation even when these were not used as statistical variables as such (e.g. employment). Also reweighting made extensive use of auxiliary information from the registers (calibration to margins). For the HFCS, some weights were scaled down to mitigate the effect of outliers in wealth variables; otherwise, weights are the same as in EU-SILC.

Appendix 1b

Estimation of individual pension plans using longitudinal registers

Individual pension plans for the 2009 survey were estimated based on the Finnish individual tax register using the so-called perpetual inventory method. Individual pension plan contributions (investments) and, respectively, pension payments received from 1990 onwards, are available in the tax register. The longitudinal tax register data were record linked to the sample of the IDS/SILC (individuals) through personal identity codes. The values of individual pension plans in 2009 were derived cumulatively from the flow data by calculating a yield for the annual net investments (contributions/payments received) as interest on interest.

In order to take into account the yield and expense structures typical of different insurance types, the plans were roughly divided into all those taken out before 1999, which were regarded as being tied to the base rate, and those taken out after 1998, which were regarded as investment linked policies. Division is based on the statistics on the sales of new life insurance policies compiled by the Federation of Finnish Financial Services, according to which the sales of new investment linked individual pension plans exceeded that of plans tied to the base rate in 1999. As the annual yield of insurance plans tied to the base rate, a base rate of 2.5 per cent, and additionally, a typical customer discount of two per cent, was applied. For investment-linked policies, the nominal annual yields of pension investment plans in the private sector were used.

The expense loading was taken into account regarding both the premium and the insurance savings. The expense loading, or plan management expenses, are the part of the insurance premium that is intended to cover the business expenses of the insurance company. The expense loadings of plans vary depending on plan type, sum of premium and the company. The expense loading is covered from either insurance premiums or insurance savings, or both. The manner in which the expenses are charged also differs.The deduction made in the insurance premium was, regardless of the policy type, taken into account by deducting five per cent from annual insurance premiums up to EUR 3,000. In addition, 0.4 per cent was deducted as expenses from the annual yield of insurance plans tied to the base rate. For investment-linked insurance plans, 0.8 per cent was deducted as expenses.

Compared to the previous Statistics Finland's wealth survey of 2004, the proportionate share of households with pension plans doubled from 12 per cent to 24 per cent. In view of the increase in demand for pension plans in Finland, this result is credible. The reliability of the results is supported by a comparison with other available statistical information. According to statistics on savings, credit uptake and payment methods of households by the Federation of Finnish Financial Services, 12 per cent of individuals has voluntary pension plans. The estimated share of individuals in the survey is slightly higher, 15 per cent. Additionally, the total of insurance plans calculated in the 2009 survey is fairly close to the aggregates put forward by the Federation of Finnish Financial Services in their statistics on the sales of new life insurance policies. The average investment value of those who have pension plans, on the other hand, decreased by one quarter, which reflects the prevalence of small sum investments in particular among such groups as young people.

Appendix 2

Capacity to face unexpected financial expenses according to household respondent ("Subjective liquid asset poverty") and objective asset-based poverty measures, % of individuals.

	Capacity to face unexpected expense of EUR 1,000	Objective asset-based poverty measures with different wealth concepts and thresholds								
		Liquid financial wealth			Excluding home equity			All net worth		
		1 month	3 months	6 months	1 month	3 months	6 months	1 month	3 months	6 months
All	28.1	25.9	46.1	61.6	23.5	32.4	43.2	16.2	20.5	24.7
Socioeconomic status										
Farmers	10.2	11.6	28.6	47.1	21.1	23.7	27.1	4.0	4.0	4.1
Self-employed	13.9	19.2	40.0	55.6	23.5	28.8	34.9	14.1	16.3	17.5
Employees, upper level	12.4	16.5	33.4	49.1	17.9	23.1	30.7	12.2	14.7	17.7
Employees, lower level	27.0	24.5	46.8	64.1	24.9	33.6	45.1	19.7	23.7	28.2
Manual workers	30.9	29.8	52.7	68.5	27.8	36.0	46.9	19.7	24.0	28.9
Students	41.4	35.1	57.9	72.4	33.1	44.2	56.2	25.3	32.4	37.9
Pensioners	23.2	16.6	32.0	48.1	15.0	23.4	34.6	7.9	11.0	14.5
Others	31.7	33.4	56.9	71.5	26.3	36.7	49.1	18.0	22.8	27.2
Unemployed	64.1	48.7	67.4	78.0	35.8	52.6	64.0	29.9	41.1	48.8
Age										
0–24	35.4	34.3	57.4	72.1	28.5	39.3	51.8	20.9	26.7	31.8
25–34	34.6	35.8	58.6	74.9	35.2	44.0	56.8	33.9	40.5	47.7
35–44	26.9	26.5	49.2	64.5	23.6	32.1	42.3	14.6	18.2	22.0
45–54	27.0	23.3	43.1	59.7	23.2	31.9	41.0	12.6	15.8	19.1
55–64	20.5	14.8	31.5	47.7	15.4	21.7	29.0	7.9	10.5	12.8
65–	18.9	15.6	30.0	45.1	13.3	21.7	33.1	6.3	9.1	12.1
Household size										
1	41.3	32.2	50.2	64.6	28.2	40.7	53.7	22.5	29.8	36.6
2	21.3	18.6	35.9	51.4	20.0	27.3	35.9	15.0	18.8	22.5
3	26.5	24.4	47.4	63.7	23.7	31.7	41.6	16.8	21.3	24.8
4	25.5	27.2	51.4	67.7	21.9	29.6	41.5	13.6	16.2	20.1
5	30.3	33.4	53.0	70.3	26.8	35.6	46.9	13.4	15.8	18.0
6+	34.5	39.3	64.5	74.5	29.0	42.1	58.4	12.9	17.7	21.7

Source: Wealth Survey 2009 / SILC 2010

Appendix 3

Income poor and asset poor, objective and subjective (capacity to face unexpected expenses) asset-based measures, % of individuals.

	Income poor		Twice-poor (income and asset poor)				
	With IR	Cash	With IR, net worth excl. home equity, 3 months	Cash, net worth excl. home equity, 3 months	With IR, subjective asset poverty	With IR, liquid assets, 1 month	With IR, liquid assets, 3 months
All	13.1	13.1	8.2	7.0	8.7	7.1	9.8
Socioeconomic status							
Farmers	11.4	17.9	3.3	5.1	1.6	3.2	5.4
Self-employed	11.8	13.4	5.5	5.6	4.3	4.0	7.6
Employees, upper level	0.7	0.8	0.5	0.6	0.4	0.3	0.5
Employees, lower level	3.3	2.1	2.1	1.3	2.1	2.0	2.5
Manual workers	4.9	3.1	3.2	1.8	3.1	2.7	3.6
Students	32.2	29.8	23.0	19.8	22.2	18.0	24.6
Pensioners	14.0	18.3	7.5	6.8	7.9	5.8	9.3
Others	15.0	13.1	9.4	7.7	10.4	9.1	12.0
Unemployed	49.5	46.2	34.4	30.4	41.9	32.0	40.4
Age							
0–24	18.1	15.8	12.1	10.2	12.8	10.8	14.2
25–34	13.3	11.3	9.3	7.6	9.6	8.3	10.6
35–44	10.2	9.6	5.9	5.2	7.2	5.6	7.5
45–54	9.9	9.2	6.3	5.2	7.4	5.0	7.3
55–64	9.5	9.9	5.5	3.9	6.5	4.0	7.1
65–	12.2	18.6	6.3	6.5	5.2	5.3	7.9
Household size							
1	28.4	31.7	18.0	16.0	18.4	14.9	20.9
2	8.0	7.5	5.2	4.0	5.5	4.0	5.8
3	8.6	8.3	5.5	4.5	6.5	5.1	7.0
4	9.2	8.3	4.5	4.0	6.1	4.7	6.6
5	11.4	7.9	8.3	5.3	8.6	7.0	8.6
6+	21.1	22.8	13.1	14.9	10.4	13.1	17.6

Note: IR=net imputed rents of owner-occupiers. Cash = cash and near cash income.

Source: Wealth Survey 2009 / SILC 2010

Appendix 4

Poverty profiles based on wealth-augmented poverty measures, relative risk profiles, % in the subroup / % in the total population, Finland 2009.

	Income poor	Asset poor	Income and asset poor	Vulnerable	AROPE	W-AROPE 1 (3 dimensions, k=1)	W-AROPE 2 (4 dimensions, k=2)
All	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Socio-economic status							
Self-employed	0.9	0.8	0.6	0.9	0.7	0.4	0.5
Employed	0.2	1.0	0.2	1.2	0.2	0.2	0.2
Students	2.5	1.4	2.8	0.9	2.2	2.3	2.4
Retired	1.1	0.7	0.9	0.7	1.2	1.1	1.0
Unemployed	3.8	1.7	4.2	0.8	3.8	4.6	4.7
Others	1.1	1.1	1.1	1.1	1.1	1.1	1.2
Age							
0–24	1.4	1.2	1.5	1.1	1.3	1.3	1.4
25–34	1.0	1.4	1.1	1.4	1.0	1.0	1.1
35–44	0.8	1.0	0.7	1.1	0.8	0.9	0.9
45–54	0.8	1.0	0.8	1.1	0.9	1.1	1.0
55–64	0.7	0.7	0.7	0.7	1.0	1.0	0.8
65–	0.9	0.7	0.8	0.6	0.8	0.6	0.6
Income level (% of median)							
– 40	7.6	2.2	8.5	0.0	6.1	6.5	7.6
0– 50	7.6	2.1	8.2	0.0	6.1	6.3	7.4
0– 60	7.6	1.8	7.1	0.0	6.1	5.2	6.0
0– 70	0.0	1.4	0.0	1.9	0.9	1.1	0.7
70–100	0.0	1.1	0.0	1.4	0.3	0.4	0.2
100–150	0.0	0.8	0.0	1.1	0.1	0.1	0.0
150–200	0.0	0.4	0.0	0.6	0.0	0.1	0.0
200–	0.0	0.2	0.0	0.3	0.1	0.1	0.0

Source: Wealth Survey 2009 / SILC 2010

Tilastokeskus
Statistikcentralen
Statistics Finland